Available online at www.sciencedirect.com

ScienceDirect

Procedia - Social and Behavioral Sciences 111 (2014) 829 – 838

Procedia
Social and Behavioral Sciences

EWGT2013 – 16th Meeting of the EURO Working Group on Transportation

Consistently estimating link speed using sparse GPS data with measured errors

Masoud Fadaei Oshyaniv[a,]*, Marcus Sundberg[a], Anders Karlström[a]

[a]*KTH Royal Institute of Technology, Transport and Location Analysis, SE-100 44 Stockholm, Sweden*

Abstract

Data sources using new technology such as the Geographical Positioning System are increasingly available. In many different applications, it is important to predict the average speed on all the links in a network. The purpose of this study is to estimate the link speed in a network using sparse GPS data set. Average speed is consistently estimated using Indirect Inference approach. In the end, the Monte Carlo evidence is provided to show that the results are consistent with parameter estimates.

Keywords: Travel time; Sparse GPS data; Indirect inference; Map matching; route choice.

1. Introduction

Travel time is a critical aspect of all trips which is usually considered by people in their route choice. It provides a key aspect in transportation planning and appraisal, and to be able to accurately measure travel time is of paramount importance. For instance, travel time is related to other key factors such as congestion and pollution, and also has a significant impact in social cost benefit analysis, both directly and indirectly.

With increasing availability of data using new technology such as the Geographical Positioning System, new methods and algorithms are being developed that are tailor-made for the new data sources to address specific problems. For instance, to provide route guidance in real-time to emergency services vehicles to reduce the travel time, Westgate et al. (2011) develops a Bayesian model of the ambulance trips. Miller et al. (2010) used GPS data containing speed and location; whereas Westgate et al. (2011) used data including speed, location and timestamp as well. Furthermore, several studies had recently been done to estimate travel time or speed value for arterial road segments in a network based on sparse GPS data (e.g. Hofleitner et al., 2012; Jenelius & Koutsopoulos,

* Corresponding author. Tel.: +46-8790-6897; fax: +46-8790-7002.
E-mail address: masoud.fadaei@abe.kth.se

1877-0428
doi:10.1016/j.sbspro.2014.01.117

2013). In this paper we deal with data containing just location and timestamps to estimate the values of link speeds.

Generally, GPS data in the literature is classified into low and high frequencies. If traversed paths between all two consecutive GPS points can be accurately detected, this data is defined as high frequency; otherwise it would be low frequency. On the other hand, low frequency data are also used for various reasons, including privacy issues, data collecting and storage costs. When the data is collected with low frequency, the path traversed between two consecutive GPS data points are not always known with certainty. Although most studies on travel time estimation have been done on high frequency data (e.g., Work et al., 2008 and Zou et al., 2005), examples with low frequency data include Jenelius and Koutsopoulos (2013). Another problem that should be acknowledged, in particular with low frequency data, is the measurement errors associated with GPS data. Chen et al. (2005) report 27 meters as average positioning errors in their dense case which was part of Hong Kong network.

To make the most use of low frequency GPS data sets with measurement errors, new methods and algorithms are being developed to address specific objectives. The purpose of this study is to estimate the link speed in a network using such a sparse data set. In our previous work (Fadaei Oshyani et al., 2012) we developed a method for estimating route choice models when data is spatially and temporally sparse. This method was based on a consistent estimator proposed by Karlström et al. (2011) for estimating route choice models the complete paths are observed (high frequency data). Following the indirect inference approach used in these previous studies (Gourieroux et al., 1993, Karlström et al., 2011) we develop an estimator for link speeds, and show how we can jointly estimate both parameters of the route choice model, and link speeds.

The route choice model that this study is based on is link based with random costs, and where the path cost is additive in link costs. Such a model exhibits flexible correlation structure and thus a realistic substitution pattern (Frejinger & Bierlaire, 2007). Also, it is useful from an application point of view since it is easy to generate paths given the model and its parameters. The route choice model describes how routes are chosen in the network, which is crucial to understand when the paths are not known, using sparse GPS data. Given a route choice model, we develop an estimator for link speeds. Finally we show how the two estimators for (i) the parameters of the route choice model and (ii) link speed both can be brought together into an algorithm where both are estimated iteratively.

The paper is organized as follows. First, we state the proposed model that is used for estimating the average speed on the network links. Then we specify our indirect inference-based estimator. At the end, the iterative process of estimation will be explained in which we will estimate the average speed on the network links.

2. The Model

In this section we will introduce the data generating process that is believed to generate sets of sparse GPS data with some measurement error. Our main objective is from such a data set to infer link speeds. In our study, the travel data set consists of travel records of vehicles' locations and the associated time stamps. Each observation is associated with a trip that is represented with a set of GPS points, each of which having latitude, longitude, and time stamp. Assuming that the origin and destination of the path is known and located on nodes on the road network, we obtain partial observation of the path in between. Firstly, we do not know the path taken between the origin and destination, and, secondly, the GPS points are generated with a measurement error.

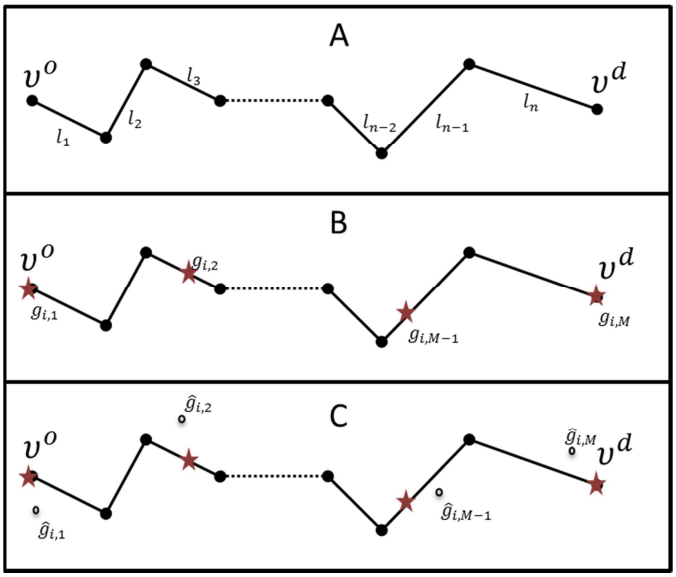

Fig. 1. (A) A path between an origin and destination is simulated by means of the route choice model. (B) Conditional on the route, a car traversing this route is simulated according to average link speeds, providing positions and time stamps. (C) The final relocation based on the GPS measurement error.

We propose a model with a probabilistic route choice component, and a GPS point simulator conditional on the route. For such model, there are different types of parameters to be specified such as behavioural parameters, sampling frequency, GPS data measurement error and average speeds on all the network links. In this paper we are primarily interested in estimating the average speeds, it is assumed that the sampling frequency and measurements error are known and the average link speeds are going to be consistently estimated by having the observations, i.e. a travel data set in sparse GPS format. Behavioral parameters will also be estimated, though they are not the main concern in this paper. There are two related and yet unobserved components associated with each trip. The first is the traversed path and the second is speed values on the links. Our model that generates sparse travel data consists of two parts. The first part is a route choice model with the parameter β. This model introduces the path traversed by travelers. The second part of the model is a simulator that generates GPS points from a given path with some known measurement error, data sampling frequency F, and average speeds on all the links of the network.

We consider individuals making route choices on a road network given their origin and destination. The network $\mathbb{N}$ in our model consists of sets of nodes (v) and links (l). A path connecting an origin node v^o to a destination node v^d is shown with a sequence of links making the path. Therefore, the indexes of these links could be used to introduce the path $\pi = \{l_1, \dots, l_n\}$. Each link has its own strictly positive cost function $c(x_l, \epsilon_{l,i}; \beta)$ depending on the vector of its specific characteristics x_l, (e.g. link travel time). The vector β is the coefficient of the link characteristics and $\epsilon_{l,i}$ is the random link cost for individual i on link l. The cost function in this paper is chosen as linear (Eq. 1).

$$c(x_l, \epsilon_{l,i}; \beta) = \beta x_l + \epsilon_{l,i}, \tag{1}$$

Based on the path definition, the cost of each path π is equal to the summation of costs for its consisting links. Therefore, the cost for individual i to pass a path π is calculated by

$$C_i(\pi) = \sum_{l\in\pi} c\left(x_l, \epsilon_{l,i}; \beta\right). \quad (2)$$

In addition, the travellers want to maximize their utilities, i.e. minimize their cost, based on their known idiosyncratic random utility $\epsilon_{l,i}$ and the link characteristics regarding their passed links; therefore, they will take the path with the minimum generalized cost in this model

$$\pi_i = \arg \min_{\pi\in\Omega(v_i^o, v_i^d)} C_i(\pi), \quad (3)$$

where $\Omega(v_i^o, v_i^d)$ denotes all possible paths connecting v_i^o (origin) and v_i^d (destination). In our case, the random part $\epsilon_{l,i}$ is assumed to follow a truncated normal distribution using only the positive values, in order to avoid negative link costs. This assumption leads to always having a positive link cost on the network; thereby satisfying the prerequisites of the Dijkstra shortest path algorithm. Therefore, in order to simulate path choices in accordance with (3), we draw individual specific random components, calculate link costs, and then apply the Dijkstra algorithm to find the individual's shortest path. The output is a path consisting of a set of links which is illustrated in Figure 1-A.

To generate a set of GPS points from a path, a virtual car is sent through the path and driven with the average speed on the corresponding links. The location of the car is determined and recorded every $1 / F$ seconds and denoted as $G_i = \{g_{i,1}, \dots, g_{i,M_i}\}$ for individual i, where M_i denotes the number of GPS points for the observation i. These points are shown in Figure 1-B with stars.

As mentioned before, GPS data is noisy, thus causing the aberration between the measured location and the true location on the map. We assume that all the GPS data collectors have a known error distribution with zero mean modelled as a two-dimensional symmetric Gaussian, in latitude and longitude. According to this assumption, the GPS points are re-located by adding a measurement error, drawn from the symmetric normal distribution with a given standard deviation. Finally we generate a series of GPS data points passed in a trip through path π as $\hat{G}_i = \{\hat{g}_{i,1}, \dots, \hat{g}_{i,M_i}\}$; where $\hat{g}_{i,1}$ is the first GPS point and $\hat{g}_{i,M_i}$ is the last one for the individual i (Fig. 1-C).

3. Method

Due to the large number of links in a typical network (in our case, 7459 links) and a relatively low number of observations by GPS, it is practically impossible to estimate the average speed on all these links due to the limitations of the data set. Thus, we categorize the links into J different classes, in our case according to their speed limits, and only J parameters need to be estimated, as explained in more details later on. Also, the sparsity of GPS data implies that the traversed path by the traveler is not known with certainty. The idea here is to use a route choice model (our previous work, Fadaei Oshyani et al., 2012), and generate a path according to the source and destination and the behaviour of the traveler on the network (See Fig. 1).

Although we have a model which generates GPS data sets, there is no straight forward way of setting up a likelihood function to be maximized in order to estimate average speed values. In large, this difficulty arises due to the unobserved route choice which generated the GPS points. As a remedy, we propose the use of an Indirect Inference approach for estimation, which is a simulation-based estimation technique (Gourieroux et al., 1993).

In this section we will develop an indirect inference method for estimating the values of average speed on all the road segments in the network based on the observed travel data set in sparse GPS format. Indirect inference is a simulation-based method to estimate econometric models. Thus, as an initial requirement, the model of interest should be able to simulate data for a variety of parameter values. The main characteristic of the indirect inference

method is the use of an approximate or auxiliary model in order to form a criterion function. The number of parameters for the auxiliary model has to be more or at least equal to the number of parameters in the main model. There are two requirements for choosing an auxiliary model. First, it should be easy to estimate, since we want to get help from an auxiliary model to estimate its parameters (the auxiliary parameters) and run the auxiliary model repeatedly. Second, the auxiliary model has to be flexible enough to capture the variation of the observed data. The aim of the indirect inference is to find the parameters for the main model such that the simulated and observed data look the same from the auxiliary model's point of view.

The indirect inference approach aims to estimate a model of interest, which is advantageous in practice, but it is difficult to estimate. In our case the true model is described in Section 2. This model is able to simulate a data set with given values for route choice parameter β, and average speeds on the links and given values for sampling data frequency and GPS data measurement error. To simulate a series of GPS points for a trip between a given origin and destination pair, we need to run the route choice model to simulate traversed path based on its characteristics and β; then, we take the data sampling frequency and speed values and implement the GPS point simulator in the model for the traversed path. In addition, an auxiliary model is required, this model may thus be misspecified but easy to estimate and powerful enough to capture relevant variation in the data.

In our auxiliary model, for a series of GPS data points passed in a trip by individual i denoted by $\hat{G}_i = \{\hat{g}_{i,1}, \dots, \hat{g}_{i,M_i}\}$, there are a series of time stamps $\{t_{i,1}, \dots, t_{i,M_i}\}$ where $t_{i,m}$ denotes the time that the traveler passed the GPS point $\hat{g}_{i,\mathrm{m}}$. First, through the auxiliary model, we may find the most likely path which was traversed by the traveler while passing the GPS points $\hat{G}_i$. In the literature, there are several methods for mapping GPS data to digital networks and estimating the traversed path. These methods are generally named as map-matching techniques. In our case, we apply our own personalized map-matching method that will be explained later on in this section. We here assume to detect the actually traversed path π with the help of our map-matching method. It is notable that our map-matching method may be misspecified. However, as it is a part of the auxiliary model, misspecification and mismatched paths will not compromise the consistency of our final estimates, this is a strength of the indirect inference approach.

Then, we define the closest point on the matched path (output of map-matching method) to the GPS point $\hat{g}_{i,\mathrm{m}}$ as its inferred matched location $\acute{g}_{i,m}$. $\acute{G}_i = \{\acute{g}_{i,1}, \dots, \acute{g}_{i,M_i}\}$ denotes the set of inferred matched points on the path π regarding GPS points $\hat{G}_i$. Since we assume the origin and destination of the trips to be known, we map the first and last GPS points to them.

Then we again send our virtual car through the path and let it drive based on the auxiliary parameters as the values for average speeds. The car will pass point $\acute{g}_{i,m}$ at a specific time $\acute{t}_{i,m}$. Since we want to estimate the values of average speeds on the links $\big(S(l)\big)$, parameter $Z(l)$ is defined as our auxiliary speed parameter for the link l in the network in which $Z(l)$ is related to its speed $S(l)$. The parameters of the auxiliary model will be estimated to minimize the summation of squared differences between $t_{i,m}$ and $\acute{t}_{i,m}$.

As mentioned estimating the average speed is costly for all the existing links in a network, and could be practically impossible given a limited data set. Therefore, in this study we categorize all the links in the given network into J classes based on the speed limits and lengths associated to each link. In such conditions, the average speed for link l is equal to S^j if link l belongs to class j. This classification decreases the number of parameters, the average speed on all the network links, to only J ones. Furthermore, the number of unknown

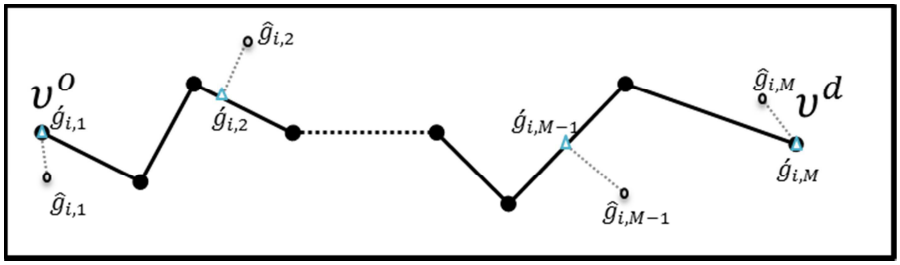

Fig. 2. Provided GPS points and a matched path, the closest location on the path is determined.

parameters in the auxiliary model should be equal to or more than the number of parameters in the main model (Smith, 2008); therefore, the vector of $Z = \{Z^j\}_{j=1}^{J}$ is defined for the same number of elements as $S = \{S^j\}_{j=1}^{J}$ has.

In theory, there is a correspondence between the parameters of the true model and the auxiliary parameters which is revealed through a smooth binding function $Z(S)$. This function should be explored by simulations. Since the average speeds are the only parameters of interest for us, we initially fix the value β and draw a set of random structural speed parameters $\{ S_k^j\}_{j=1}^{J}, k = 1, \dots, K$ from specified domains of interest $\mathcal{D}^j$. Then K different GPS data sets will be simulated using our true model. These simulated GPS data sets are converted into matched paths by applying our map-matching method. The set of matched paths are denoted by $\tilde{y}(S_k)$, where $S_k = \{ S_k^1, \dots, S_k^J\}$ is the vector of speed values. Then we estimate the vector of auxiliary parameters corresponding to each specific data set simulated by the same parameters $\tilde{Z}_k(S_k)$.

A smooth binding function is estimated by local OLS, based on K different given values of S_k and their corresponding $\tilde{Z}_k(S_k)$. This smooth binding function is denoted by $\tilde{Z}(S)$.

The purpose of the indirect inference method is to select the parameters of the true model such that the simulated and observed data look the same from the auxiliary model's perspective. For estimating speeds, a least square loss function is used, measuring the distance between observed time stamps ($t_{i,m}$) and corresponding traversed time ($\dot{t}_{i,m}$) for inferred matched points, based on the auxiliary model. Thus, given observed series of GPS $i = 1, \dots, I$, where each series consists of a set of M_i GPS points, we estimate

$$\hat{S} = \arg\min_S \sum_i \sum_{m \in M_i} (t_{i,m} - \dot{t}_{i,m}(\tilde{Z}(S)))^2. \quad (4)$$

One of the main characteristics to consider in the route choice model is travel time. Typically we want to determine the effect of travel time on route choice, by estimating the corresponding route choice parameter. As link travel times have to be inferred from link lengths and estimated link speeds, a correlation is introduced between the estimates of the route choice parameter (β) and the link speeds $S(l)$. Therefore, we propose a method to simultaneously estimate these two correlated parameters. In other words, a robust estimate for β could lead to accurately estimate the speed. Thus, we present an iterative method consisting of two sub-estimators. The first sub-estimator takes a given value for β and estimates the speed values and the second one estimates β given the speeds for the links in the network. The second sub-estimator for β is described in detail in Fadaei Oshyani et al. (2012). In other words, we assign an initial value to β and estimate speed values in the first sub-estimator. Then we take these estimated speed values and apply the estimator in the second sub-estimator to re-estimate the route choice parameter (β) and use the new value for β to run the first sub-estimator again. This process will continue until we reach the final estimate values for speeds. Both sub-estimators are constructed based on the indirect inference approach.

3.1. Map matching

The locations reported by the GPS do not accurately match to the network of digital maps; thus, we need to apply some methods to map the reported points onto the network. In addition, our map-matching method should detect the most likely traversed path for each series of travel points. For this purpose, there are several map-matching methods, in the literature. Quddus et al. (2007) analyzed 35 map-matching methods that had been presented during 1989- 2006. Lou et al. (2009) mentioned that most of the existing map-matching methods work with a high sampling rate from 6 to 2 points per minute, and do not operate well for low-rate sampled points.

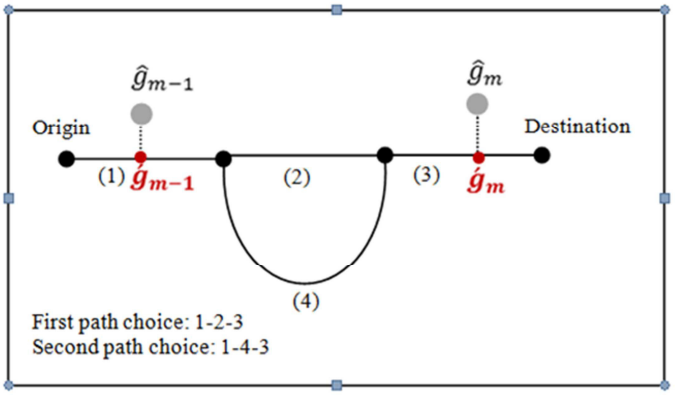

Fig. 3. Provided sparse GPS data, alternative routes may be consistent with reported locations of GPS points. By also using the time stamps we detect a probable route.

Our own map-matching technique works in the following way. As mentioned the origin and destination of the trips are assumed to be known. Thus, in the first part, many of the possible path choices connecting the given origin and destination pairs are identified. Then, the summation of distances between each possible path and the set of GPS points associated with the given trip is calculated. Finally, the path with the minimum distance is introduced as the matched path.

Although the method presented in the first part works well for the GPS data with high frequency and low measurement errors, there is a difficulty for its application in our case with sparse data. Due to sparseness in our dataset, in practice, the first part of our map matching method usually returns more than one choice as the matched path. In other words, there are usually a number of sub-paths connecting two consecutive points resulting in having different paths with the same total distance to the set of GPS points. (Fig. 3)

In order to deal with this problem, we introduce the third step in which the aim is to find the best matched path from these different output paths from the second step. A solution to deal with this problem is to find the actual locations of the GPS points on a path based on the point time stamps and speeds on the links; then, calculating the distances between these locations and their reported positions in the dataset. The choice with minimum distance is detected as a final matched path. The actual location for a GPS point on a given path is defined as the location on the path that the traveler would be there if he/she took that path.

Assume $g_{i,m}$ represents the actual position of the GPS point $\hat{g}_{i,m}$on a given path. To find the actual locations of the GPS points for the given path, we send a virtual car through the path and let it drive based on assigned speeds to the links belonging to the path. After traveling for $t_{i,m}$ as the timestamp associated with $\hat{g}_{i,m}$, the actual location of the GPS point will be detected as $g_{i,m}$.

Since the information regarding the links speed values is still unknown, we substitute the link speeds in our data set with the speed limits as the average speed values on the links. Although speed limit is not an accurate estimate for the average speed, it could be practically helpful in the indirect inference approach. In other word, since we apply our map-matching method for both real and simulated data in the same way, the error introduced in map-matching will be corrected for by the indirect inference based estimator.

4. Case study

We used the transportation network in Borlänge city, Sweden, representing a directed graph containing 3077 nodes and 7459 links (Frejinger & Bierlaire, 2007). All the links in the network were divided into four speed-

classes ($S^1, \dots, S^4$) based on their speed limits and lengths. As the assigned speed limits to the links for the Borlänge network vary from 5kph to 100kph, the speed-classes must be formed to cover all the links in the network. Furthermore, we assumed that the speed on a link is fixed and equal to the average speed. For this case study, link travel time was considered as the parameter of interest in the route choice model. The link travel times are not given by data, but have to be inferred from reported link lengths ($L(l)$) and the estimated link speeds ($S(l)$), thus we estimate the route choice model based on the link travel time characteristic ($x_l = L(l)/S(l)$)

To verify the accuracy of our proposed method, we first simulate a dataset based on our route choice model and then estimate parameters of interest using the proposed method in section 3. Given the network and the route choice model, a value is assigned to β (say $\beta = 2$) and four values to average speeds for link-classes (say $S^1 = 20kp/h, S^2 = 32kp/h, S^3 = 55kp/h, S^4 = 70kp/h$); then, datasets are simulated. We simulated data for $N = 3000$ trips. For each trip, while traveling along the path, the location of the virtual car was stored every 60 seconds, considering known GPS measurement error. Finally, our indirect inference-based estimator was applied.

In this paper, since the "true" speed values are known, we choose an appropriate interval for generating the binding function, it is $\mathcal{D}^1 = (5;\ 30), \mathcal{D}^2 = (30;\ 40), \mathcal{D}^3 = (40;\ 60), \mathcal{D}^4 = (60;\ 80)$. The indirect inference based estimation were done with a binding functions which themself are estimated using 10 sample points drawn from $\mathcal{D}^j$ for speed classes. For each such sample point the auxiliary parameters are estimated based on $N = 3000$ simulated paths. GPS related spatial error is introduced following a normal distribution $N(0;\ \delta)$. We examined whether the estimated parameters are consistent with the assigned values to the model for simulating datasets.

4.1. Results

All Monte Carlo statistics are calculated based on 10 independent estimations of the parameters. That is, we create ten independent sets of "observed" GPS data and then we apply our iterative Indirect Inference-based estimator once to each of these data sets. In Table 1 we report the Monte Carlo evidence of the average speed estimates for three GPS sampling interval. According to the reported results, estimated values for all the link classes are quite precise in the second iteration. For each of the estimates, the true data generating speed parameters falls within the 95% confidence interval of the estimated parameters. Thus the true values of speeds cannot be rejected.

Table 1. MONTE CARLO EVIDENCE: ESTIMATES OF AVERAGE SPEED VALUES FOR DIFFERENT GPS SAMPLING TIMES, AND TWO ITERATIONS.

GPS sampling time (sec)		S^1		S^2		S^3		S^4	
		Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.
60	First iteration	18.11	0.99	29.76	1.53	55.88	0.77	67.86	0.25
	Second iteration	19.65	0.38	32.34	0.89	55.69	0.54	69.81	0.42
90	First iteration	18.10	0.67	30.43	1.01	55.36	0.53	67.58	0.29
	Second iteration	19.75	0.47	32.45	1.09	55.23	0.60	69.63	0.40
120	First iteration	20.12	2.24	25.86	5.14	55.92	2.30	67.43	0.35
	Second iteration	22.08	0.75	29.06	0.99	55.47	1.40	69.35	0.32

Although the main purpose of the method is to estimate average speed values, our iterative estimator yet returned an accurate estimate for the route choice parameter. Table 2 shows the Monte Carlo evidence of β for the three GPS sampling interval. The estimate for the route choice parameter is precise for all data sampling values.

Table 2. MONTE CARLO EVIDENCE: ESTIMATES OF β FOR DIFFERENT GPS SAMPLING TIMES, AND TWO ITERATIONS

GPS sampling time (sec)		β	
		Mean	Std.
60	First iteration	1.946	0.037
	Second iteration	2.069	0.029
90	First iteration	2.021	0.043
	Second iteration	2.019	0.026
120	First iteration	1.972	0.076
	Second iteration	2.042	0.068

5. Conclusion

In this paper we have proposed a method to estimate the average speed values on all the links in a network using GPS data sampled with a low frequency. Travel time is calculable based on the estimated speeds for all the given trips. As mentioned travel time is a critical aspect of all trips which is usually considered by people in their route choice. In addition, it provides a key aspect in transportation planning and appraisal, and to be able to accurately measure travel time is of paramount importance. When the travel data is collected with low frequency, it is unknown which path has been traversed between the GPS data points. Moreover, GPS data has measurements error. These characteristics may introduce bias into the estimates governing route choice behavior.

We have designed an iterative method for the two estimators for the parameters of the route choice model and link speed both which have been brought together into an algorithm. First, the links are classified based on their speed limits into a number of classes. Then, the average speed of each class is consistently estimated. We have applied a route choice model which is link based with random costs, and where the path cost is additive in link costs. The route choice model describes how routes are chosen in the network, which is crucial to understand when the paths are not known, using sparse GPS data.

The main conclusion is that indirect inference is a useful option for estimating speed values on all the links in a network. Our indirect inference based method can be used for estimating speeds using low frequency GPS sampling data with measurement errors. The Monte Carlo evidence shows that, applying the indirect inference approach to speed estimation is a worthwhile solution.

References

Ben-Akiva, M., & Bierlaire, M. (1999). Discrete choice methods and their applications to short term travel decisions, in R. Hall (Ed.). *Handbook of Transportation Science* (pp. 534). Kluwer, Dordrecht, The Netherlands.

Cascetta, E., Nuzzolo, A., Russo, F., & Vitetta, A., (1996). A modified logit route choice model overcoming path overlapping problems. Specification and some calibration results for interurban networks, in J. B. Lesort (Ed.). *Proceedings of the 13th International Symposium on Transportation and Traffic Theory*, Lyon, France.

Chen, W., Li, Z., Yu, M., & Chen, Y. (2005). Effects of sensor errors on the performance of map matching. *The Journal of Navigation, 58,* 273 - 282.

Fadaei Oshyani, M.; Sundberg, M., & Karlstrom, A., (2012). Estimating flexible route choice models using sparse data, *Intelligent Transportation Systems (ITSC), 15th International IEEE Conference on* , (pp.1215,1220).

Frejinger, E., & Bierlaire, M., 2007. Capturing correlation with subnetworks in route choice models. *Transportation Research Part B: Methodological 41(3)*: 363–378.

Gourieroux, C., Monfort, A., & Renault, E., (1993). Indirect inference. *Journal of Applied Econometrics,* 8(S1):S85–S118.

Hofleitner, A., Herring, R., & Bayen, A., (2012) Arterial travel time forecast with streaming data: A hybrid approach of flow modeling and machine learning, *Transportation Research Part B: Methodological, Volume 46, Issue 9*, Pages 1097-1122

Jenelius, E., & Koutsopoulos, H. N. (2013). Travel time estimation for urban road networks using low frequency probe vehicle data. *Transport Research Part B 53*, 64 - 81.

Karlström. A., Sundberg, M., & Wang, Q., (2011). Consistently estimating flexible route choice models using an MNL lens. *International Choice Modelling Conference*, Leeds, UK.

Lou, Y., Zhang, C., Zheng, Y., Xie, X., Wang, W., & Huang, Y., (2009). Mapmatching for low-sampling-rate GPS trajectories. *In Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (GIS '09).* ACM, New York, NY, USA, 352- 361.

Miller, J., Kim, S., Ali, M., & Menard, T. (2010). Determining time to traverse road sections based on mapping discrete GPS vehicle data to continuous flows*. IEEE Intelligent Vehicles Symposium*, (pp. 615–620).

Quddus, M.A., Ochieng, W.Y., & Noland, R.B., (2007). Current map-matching algorithms for transport applications: State-of-the art and future research directions, *Transportation Research Part C: Emerging Technologies, Volume 15, Issue 5*, 312-328.

Smith, Anthony A., Jr., indirect inference, (2008). *The New Palgrave Dictionary of Economics*, Eds. Steven N. Durlauf and Lawrence E. Blume, Palgrave Macmillan.

Westgate, B. S., Woodard, D. B., Matteson, D. S., & Henderson, S. G. (2011). Travel time estimation for ambulances using Bayesian data augmentation. Submitted, *Journal of the American Statistical Association*.

Work, D. B., Tossavainen, O.-P., Blandin, S., Bayen, A. M., Iwuchukwu, T., & Tracton, K. (2008) An ensemble Kalman filtering approach to highway traffic estimation using GPS enabled mobile devices. *Proceedings of the 47th IEEE Conference on Decision and Control*, (pp. 5062-5068).

Zou, L., Xu, J.-M., & Zhu, L.-X. (2005) Arterial speed studies with taxi equipped with global positioning receivers as probe vehicle. *Proceedings of the 2005 International Conference on Wireless Communications, Networking and Mobile Computing*, (pp. 1343-1347).